

Methodology and Standards

1.0 Methodology and Standards Program Area Overview

The Methodology and Standards program area includes three major components: Computer-Assisted Survey Information Collection technology research, the Integrated Statistical Laboratory, and the Administrative Records Research System.

The Computer-Assisted Survey Information Collection (CASIC) plan improves how the U.S. Census Bureau designs, conducts, and manages surveys and censuses, using a set of automated tools to improve data quality, finish reports faster, and increase cost-effectiveness. In addition, the Computer-Assisted Survey Research Office (CASRO), which oversees CASIC, will continue to focus on process improvement. CASRO has established a team that uses process analysis, re-design expertise, and technology to improve U.S. Census Bureau processes. CASRO enhances current systems, implements new appropriate new production technologies, and researches and develops prototypes for emerging technologies that may apply to U.S. Census Bureau processes. CASRO needs hardware, software, and contractor services to support these tasks.

The Integrated Statistical Laboratory (ISL) conducts research on improving statistical analysis, IT processing, and decision-making. ISL gives researchers the power and flexibility to utilize multiple operating

systems, many users, and large files and databases. To continue our ISL work, we will need upgrades and replacements. We are upgrading our personal computers to the Personal Computer Management Acquisition Plan's standard; our goal is to upgrade one-third of our stock each year. As the Sun SPARC servers age, we upgrade and/or replace them with new servers with expanded disk storage capabilities; because of the expanded capacity, we usually need fewer servers. The increased disk storage also requires enhanced tape backup services.

The Administrative Records Research System is a computer-based system designed to expand statistical uses of administrative records to improve coverage of censuses and surveys, to reduce respondent burden and costs, and to generate new information not otherwise available. We also help the Decennial and Demographic program areas acquire administrative records. The Administrative Records Research System processing will evolve as we learn the effectiveness of different approaches to using administrative records to support surveys and censuses. This research has a potential payoff of up to \$3 billion if the U.S. Census Bureau uses administrative records as the primary data source for the 2010 Census.

1.1 Methodology and Standards Program Area Products, Services and Customers

Computer-Assisted Survey Information Collection (CASIC) products include automated tools that reduce survey cycle time, improve data quality, and reduce survey costs, as well as an architecture that integrates a variety of tools in a seamless system. U.S. Census Bureau program areas use these products for their surveys. Other data uses also benefit from CASIC. In addition, the Computer-Assisted Survey Research Office (CASRO) staff performs initial research to find new technologies that will benefit users. CASRO creates prototypes and completes pilot projects for our customers; then, before the technology goes into production, we write transition plans to develop milestones and address subjects such as risks, roles and responsibilities.

The Integrated Statistical Laboratory's product is research used throughout the U.S. Census Bureau. Our clients are the Statistical Research Division, including the Center for Survey Methods Research; the Planning, Research and Evaluation Division, which includes the Administrative Records Research Staff; and the Computer-Assisted Survey

Research Office. Our computer resources are available to all U.S. Census Bureau Directorates, and we strongly encourage joint research projects across program areas. These projects require support and cooperation from the IT Directorate's Telecommunications Office to help program areas access and transfer files over the network. Similarly, the Integrated Statistical Laboratory supports other decentralized network staffs within the U.S. Census Bureau.

The Administrative Records Research Staff produces statistical applications of administrative records such as censuses, surveys and estimates. We supply the U.S. Census Bureau with administrative records data given to us by federal, state and local government agencies. We also produce value-added administrative records data that has been edited, geocoded, and /or modeled according to customer requirements. Our customers include the Decennial, Demographic, and Economic program areas, as well as other areas within the Methodology and Standards Directorate.

1.2 Methodology and Standards Program Area IT Objectives

The Methodology and Standards program area seeks to make maximum and efficient use of its IT resources to meet programmatic objectives by:

- supporting work processes that have been simplified or otherwise redesigned to reduce costs, improve effectiveness, and make maximum use of commercial, off-the-shelf technology;
- producing timely, relevant, and quality products to meet customer's needs;
- develop innovative solutions to methodological and statistical problems to maintain and enhance the U.S. Census Bureau's position as the premiere statistical agency;
- develop approaches to minimize respondent burden and encourage respondent participation;
- develop a diverse workforce that is widely recognized for its excellence in survey and statistical research, methodology, computational expertise, and technological innovation;
- augment existing corporate standards to ensure quality data products;
- attain quality products through re-engineering processes; and
- develop prototype system for a unified metadata, accessed electronically.

2.0 Methodology and Standards Program Area IT Support

The Methodology and Standards program area leads the way in continuously improving the accuracy, cost-effectiveness, timeliness and relevance of U.S. Census Bureau products. We contribute to both the theory and practice of demographic and economic measurement through statistical, methodological and technological research.

Our work directly supports Goal 3 from the U.S. Census Bureau's 1999 Strategic IT

Plan: improve our information systems' ability to support our business processes. Furthermore, the projects we describe in the following pages help to implement Strategies Five (improve and expand electronic survey data collection, capture, and business systems) and Six (improve electronic information dissemination to the U.S. Census Bureau's customers).

Supporting a "Digital" Department of Commerce

Virtually everything that the Computer-Assisted Survey Research Office does supports the Department of Commerce's goal of moving away from paper processes to more efficient electronic processes. The following are some examples:

Computerized Self-Administered Questionnaire: this allows businesses to respond to U.S. Census Bureau surveys and censuses using their computers; currently, we distribute electronic questionnaires on diskettes for several economic surveys and censuses.

Computer-Assisted Interviewing processes: these include Computer-Assisted Telephone Interviewing (CATI), -Personal Interviewing (CAPI), and -Data Entry. We

are enhancing CAPI and CATI data collection tools and are moving a majority of demographic surveys to a CAPI/CATI environment.

Imaging: the U.S. Census Bureau will use this system, currently in development, for mailout/mailback surveys. Surveys that might use imaging include the American Community Survey and the Manufacturing, Shipment Orders, and Inventory (M3) survey.

Paperless Fax Image Reporting System: this has both outbound (i.e., faxing replacement forms to respondents) and inbound (i.e., receiving completed forms from respondents) components.

2.1 IT Systems Description

2.1.1 Detailed Description of Computer-Assisted Survey Information Collection (CASIC)

The U.S. Census Bureau established the Computer-Assisted Survey Research Office (CASRO) to investigate and evaluate technology advances in electronic information collection and provide the stimulus and infrastructure to support U.S. Census Bureau programs. CASRO's scope includes the integrated data collection, capture, cleaning, and post-collection processing for mandated U.S. Census Bureau surveys and censuses as well as reimbursable surveys conducted for external customers. These statistics are used to produce government indexes, leading indicators that track how well our nation's economy is doing; some examples are:

- employment statistics and producer price indexes for the Bureau of Labor Statistics;
- gross domestic products for the Bureau of Economic Analysis; and
- industrial production and capacity use for the Federal Reserve Board.

Using CASIC tools to integrate data collection, capture, cleaning and post-collection processing potentially affects all U.S. Census Bureau business processes. CASRO's staff (with support from the IT Directorate) is working with other program areas to examine and possibly re-engineer these business processes to use an integrated toolbox approach. CASRO's approach calls for re-engineering data collection, post-collection processing, and dissemination.

Data Collection

Data collection consists of the following processes:

- **Project Initiation and Management;**
- **Universe Creation and Sample Selection;**
- **Preparation for Data Collection; and**
- **Data Collection and Entry.**

CASRO's staff is working with the program areas to develop requirements and evaluate vendor software to accommodate some of these functions.

Project Initiation and Management: this is the process for formulating data collection methodology, cost estimation, clearance process, procedure and materials development as well as monitoring of costs, quality, and progress.

Historically, the resources needed to initiate and manage a U.S. Census Bureau data collection effort have varied widely and were generally approached on an ad-hoc, custom basis. A variety of commercial off-the-shelf management information systems software was normally used, but no standard progress reporting system existed. Payroll information from field representatives was collected through the mail. Managing case assignments and transmissions was also customized, with varying degrees of automation and controls. Standards were rarely used and, while a phased lifecycle approach to systems design was encouraged, the commitment of resources and standard tools to accomplish this did not exist.

CASIC is working to change that. The CASIC plan for project management includes sharing a common management information system, using standards in system design and maintenance, a centralized repository for data and metadata, and a telecommunications system that links all the parts together.

Universe Creation and Sample Selection: this involves some stratification, listing of units missed in universe creation, updates for the new units identified and application of selection criteria to the universe.

Creation of a universe for demographic and economic surveys entails using the:

- household and establishment data collected in the censuses,
- new construction data collected for demographic surveys, and
- business starts and terminations collected for economic surveys.

The CASIC plan includes tools such as Standardized Technologies-Assisted Mail Processing, Computer-Assisted Telephone Interviewing, Computer-Assisted Personal Interviewing, and Computer-Assisted Data Entry; these tools enhance census processes and subsequent updates and will, in turn, ensure speedier and more accurate universe definition. For example, field representatives use Computer-Assisted Personal Interviewing (CAPI) to list new construction units, which in turn facilitate sampling from this frame.

Currently, sample selection for surveys is largely automated. However, three of the four demographic sampling frames require a significant amount of clerical preparation, including the handling of paper maps. The CASIC plan includes using CAPI for

sampling and reduces risks and staff associated with the clerical listing procedure. The U.S. Census Bureau will be able to build in sampling intervals and then apply them accurately and uniformly as Regional Office staff share listings between surveys. Eventually, with a Graphical User Interface on CAPI, maps generated from the U.S. Census Bureau's Topologically Integrated Geographic Encoding and Referencing system can be displayed on portable computers. CASRO was involved in the planning and development of CAPI as a tool to compile, update, and/or correct address lists using automated instruments; we will eventually integrate reference maps with the listing instruments.

Preparing for Data Collection: this involves printing a questionnaire or programming a computer-assisted data collection instrument, making assignments or preparing cases for interview or mailout, transmitting questionnaires and materials to a field representative or to a respondent, or preparing instructional materials for and training field representatives.

Preparing for data collection combines processes supporting the CASIC plan of integrated and modularized systems. Forms preparation, the first of the two main preparatory processes, follows different paths depending on whether the U.S. Census Bureau will use paper documents or automated instruments to collect the data. Case management, the other main process, can be independent from the data collection instrument.

Traditionally, designing and generating paper forms involved preparing a camera copy and printing specifications for a contractor and monitoring the contract, receipt, storage and distribution of forms. The Standardized Technologies-Assisted Mail Processing (STAMP) project was a major

effort to modernize our survey mail preparation activities by re-engineering methodologies and applying advanced technologies. STAMP's goal was to establish a high volume, electronic printing facility operating within a centrally controlled, state-of-the-art mail processing environment. Together, these printing and processing components form a technologically advanced, fully integrated print-on-demand mailing system that will meet U.S. Census Bureau printing, publishing, and mail processing needs well into the next century.

Designing and generating computer-assisted questionnaires (instruments) involves developing specifications with the sponsoring office/agency, authoring the survey instrument, and transmitting the instrument to the field representatives or data capture staff.

The CASIC plan includes the continued move to computer-assisted questionnaires and standardizing the steps involved in setting up and authoring an automated collection instrument. The Computer-Assisted Survey Research Office will include tools to simplify setting up instruments through user interfaces.

Further automating survey instruments will reduce the risk of printing errors, reduce the turnaround time in correcting or fine-tuning the wording of the questions, and facilitate using previously reported information in the collection process itself. Paper and pencil surveys that would benefit from previous report/collection information often resort to re-asking for the information or preprinting the information (which is costly) on particular forms.

The Master Control System is a centralized control and tracking system for CASIC's data collection components. Currently, the Master Control System provides case

management functions for Computer-Assisted Telephone Interviewing (CATI), Computer-Assisted Personal Interviewing (CAPI), Computer-Assisted Data Entry, and outbound faxing. It allows cases to be transferred between CATI and CAPI. In the future, we will expand the Master Control System to include other CASIC data collection technologies and provide survey status information to survey managers.

Data Collection and Entry: this involves some type of respondent interaction that results in raw information; receipts processing, which may include check-in, screening, preliminary editing, data capture and coding; and resolving problems.

U.S. Census data collection systems combine automated and manual processes that we are enhancing by expanding CASIC. To date, we have concentrated the development of CASIC tools on data collection and capture. We can eliminate data keying from any statistical program that uses computer-assisted interviewing (CAI) technology, since CAI combines collection and entry into one step. Using CAI can also eliminate some of the preparations associated with paper forms, as well as many of the subsequent editing and corrections.

Many of the U.S. Census Bureau's statistical programs continue to use paper and pencil for data collection and keying for data entry. Collected forms generate a data entry workload of over 37 million records in a typical year at the National Processing Center and the Regional Offices.

While automating questionnaires represents a CASIC goal and effectively combines data entry with data collection, the U.S. Census Bureau's efforts in these areas since the early 1980s were uncoordinated and approached on an ad-hoc basis by survey or program.

CASIC will re-engineer the integrated data collection modules to meet the needs of all program areas.

Computer-Assisted Telephone Interviewing (CATI): currently, the U.S. Census Bureau is using two Computer-Assisted Telephone Interviewing software systems, CenCATI and MicroCATI. The Computer-Assisted Survey Research Office (CASRO) developed CenCATI to provide a single software system for all CATI surveys. CASRO based CenCATI on the Computer-Assisted Survey Methods Program at the University of California, Berkeley's Computer-Assisted Survey Execution System (CASES) software. CenCATI consists of the CASES Q authoring language with case management, supervisory, and monitoring functions. We will completely phase out MicroCATI, which is based on the QISC authoring language, at the end of calendar year 1999. We replaced MicroCATI because it did not provide the functionality our programs needed and we would have had to make a major software development effort to modernize the system. CenCATI benefited from the features available in the CASES authoring software, which improved instrument development time and allows us to request specific enhancements from the Computer-Assisted Survey Methods Program at the University of California, Berkeley.

Another Computer-Assisted Telephone Interviewing system in production is the Integrated Survey Processing Network, developed by the Economic Directorate to accommodate data entry for mixed-mode (mailout and CATI) business surveys and to integrate other survey processing functions. Currently, the Monthly Advance Retail Survey and the Monthly Wholesale Survey are using the Integrated Survey Processing Network system.

CASRO is working with the Demographic and Economic program areas as well as other U.S. Census Bureau divisions to plan and implement a systematic move from the CATI and CAPI DOS environment to a Graphical User Interface (GUI) Windows environment. As part of this process, we are working with the Computer-Assisted Survey Methods Program at the University of California, Berkeley to develop a Windows version of the CASES software. Currently, we are evaluating a beta version of CASES with the production version available later this year. The software includes a converter that will switch existing DOS instruments to Windows instruments. We are also evaluating Blaise, a Windows-based software package developed by Statistics Netherlands and used worldwide by survey organizations. The U.S. Census Bureau is using Blaise to collect failed edit follow-up data for the American Community Survey. Our initial experiences with Blaise have been positive, and we are considering adopting Blaise as an authoring language for Demographic and Economic surveys.

Our plans for the Windows environment include developing a case management and control system that would interface with multiple software packages, integrating all field operations (interviewing, listing, payroll) into one system, and establishing a CATI environment that can address both Demographic and Economic requirements. We also plan to research new capabilities for reducing respondent burden and for increasing data quality afforded by the Windows environment. Some surveys are still being conducted on paper because they would be difficult to conduct in the DOS environment; we are examining the possibility of converting these surveys to CATI and CAPI in the Windows environment.

The U.S. Census Bureau has expanded its authoring staff to accommodate the ever-increasing number of surveys to be automated. Another undertaking of this group is to standardize certain parts of instruments so that common variables and question screens can be used for introducing surveys, collecting household information, etc. Long-range plans include developing an authoring interface that would simplify the setup, coding, and documentation of automated instruments regardless of the data collection methodology they use.

Computer-Assisted Personal Interviewing (CAPI): in January 1994, the Current Population Survey was the first major survey to use CAPI. Since then, many new surveys have implemented CAPI; several major, older surveys have converted or are in the process of converting from paper and pencil data collection systems to CAPI. Additionally, the Census 2000 Accuracy and Coverage Evaluation program will use CAPI. For more specific information, please see the Decennial and Demographic program areas' sections of this Plan.

We are also keeping abreast of pen-based technologies for laptops. If the U.S. Census Bureau determines these systems feasible for production use, they would represent a new generation of CAPI capabilities, integrating data collection modules with area maps and locating procedures.

Computer-Assisted Data Entry: the CASIC plan recognizes that not all data collection efforts can be fully automated and combine data collection and capture into one process. There will still be a need for paper and pencil forms with attendant data entry systems.

Currently the National Processing Center (NPC) in Jeffersonville, Indiana performs data keying. In an average year, NPC keys over 37 million records, with these numbers rising by nearly 100 million when the Economic Census is over. Currently, NPC uses two data entry systems: DIGITAL VAX (which the U.S. Census Bureau used for the 1997 Economic Census), and networked PCs using KeyEntry III software (this replaced the TARTAN data entry system). The KeyEntry III system has the capability for both "heads down" and "heads up" keying. "Heads down" keying means that the data is keyed quickly with no concurrent editing beyond basic edits such as range checks; any coding is usually done prior to keying. "Heads up" data entry means that edits and/or coding are done at the time of keying.

The CASIC environment includes a multi-functional data capture facility at the National Processing Center. The facility consists of 250 networked microcomputer workstations connected to a centralized control system. The facility manages heads-up and heads-down data entry, Computer-Assisted Telephone Interviewing, and Touchtone Data Entry/Voice Recognition Entry. We will eventually use this facility as a receiving point for data reported by respondents using Computerized Self-Administered Questionnaires and other technologies.

Touchtone Data Entry/Voice Recognition Entry: this technology is in production and ready to accept additional work. Small vocabulary Voice Recognition Entry is incorporated into the Touchtone Data Entry system; we are still researching medium and large vocabulary capabilities.

Computerized Self-Administered Questionnaires (CSAQ): in 1993, the U.S. Census Bureau formed an interdivisional team to develop the functional requirements for Computerized Self-Administered Questionnaires (CSAQ). At the same time, we began to test this technology; the first test was for an in-house Survey of Surveys. We developed the Survey of Surveys CSAQ using the Computer-Assisted Survey Execution System authoring language, the same language used to develop our Computer-Assisted Telephone and Personal Interviewing instruments. We distributed this survey to 52 U.S. Census Bureau respondents in January 1994. Around the same time, an outside vendor with expertise in electronic publications and Electronic Data Interchange developed a CSAQ for the 1993 Company Organization Survey. They custom-coded this CSAQ in both C and CLIPPER programming languages and distributed it to 114 Company Organization Survey respondents in March and April 1994. This same vendor recreated the Company Organization Survey with minimal changes for subsequent statistical periods and additionally developed another CSAQ to be sent to a few companies in the 1994 Annual Survey of Manufactures. In the meantime, a 1994 Survey of Industrial Research and Development CSAQ was designed. It combined the menuing and electronic communications capabilities of Energy Information Administration's PC Electronic Data Reporting Option CSAQ system with an electronic instrument developed in the CASES authoring language. All of these CSAQs involved us mailing the diskettes out and the respondents mailing them back, and all but the Survey of Surveys offered the option to transmit the data back to the U.S. Census Bureau via modem.

Since these initial efforts to develop the diskette CSAQ technology, the Economic program area, working with a consultant, developed Windows CSAQs for the Survey of Industrial Research and Development and the Economic Census-Retail. These were programmed in Delphi. Additionally, more and more respondents began using CSAQ diskettes for the Company Organization Survey; soon, that survey and the Annual Survey of Manufactures CSAQs were rewritten in Delphi.

The Computer-Assisted Survey Research Office (CASRO) has developed a transition plan documenting a production environment for CSAQ. This plan describes how the Economic program area supports the development of a generalized CSAQ authoring system, plans to train in-house authors to use this system to develop future CSAQs, and integrates the software components for a survey. This plan also describes how the National Processing Center duplicates diskettes so the U.S. Census Bureau can mail them to respondents.

During the past year, CASRO and the Demographic and Economic program areas have performed research to extend CSAQ to the Internet. Using this will make this data collection method available to our respondents while reducing costs (by eliminating the need to prepare and distribute diskettes through the mail). We have been testing the distribution of both a Delphi executable (same as on diskette CSAQ) and an interactive HTML/JavaScript-based form on the Internet. Our focus has been on how to best present interactive forms on the Internet while maintaining security. Our current research plans include using digital certificates and developing a more flexible system that will have different levels of security based on the needs of each survey.

Imaging and Workflow: CASRO is developing an imaging system that the U.S. Census Bureau will use to mail and receive current surveys. CASRO is also researching workflow software for data capture and other survey processes. Our plan is to build upon the Decennial program area imaging application to meet the requirements for Demographic and Economic surveys and Economic censuses.

CASRO has developed and installed a pilot imaging, automated data capture, and workflow system. The initial program on this system is the 1998 Annual Survey of Manufactures. This system uses a Kodak 9500 scanner (as does the Decennial program area's Data Capture System 2000) to scan incoming questionnaires and correspondence. The system stores the questionnaires and correspondence as images, using Feith Document Database software from Feith Systems and Software. These images are immediately available for retrieval to analysts' desktops at the National Processing Center and at Headquarters. Microsystems Technology's Optical Character Recognition for Forms software extracts data from the questionnaires; any questionable characters are verified/corrected by a key-from-image clerk.

CASRO will undertake a detailed evaluation of this system to determine how effective the automated data capture (optical recognition with key-from-image verification) is and what are the benefits of using imaging technology to store images for future retrieval. Future studies include the inbound fax and workflow capabilities of the system. Programs expressing interest in using this system include the American Community Survey, the Jeffersonville Activity Reporting System, and the Manufacturing, Shipments Orders, and Inventory (M3) survey.

Paperless Fax Imaging Reporting System: the Paperless Fax Image Reporting System has both outbound (faxing replacement forms to respondents) and inbound (receiving completed forms from respondents via fax) component. The Technologies Management Office is already maintaining a production system for outbound faxing and CASRO has written a transition plan for this system. CASRO has also written a research plan for an inbound system and plans to write a transition plan. A system is currently being tested for M3; this system uses commercial off-the-shelf software packages for its three main software components: the fax receiving and routing software, the software used for the Correct From Image operation, and the image repository setup. It receives faxes from a sample of M3 respondents through a toll-free number, reads them using intelligent character recognition technology, routes the images to a program, and then sends the data through a load file to the production M3 database. A centralized system would allow the U.S. Census Bureau to maintain fewer toll-free numbers and reduce some duplication of staff responsibilities at Headquarters and the National Processing Center. It would absolve the individual branches from maintaining their own systems while still delivering data and check-in information to Headquarters quickly as surveys near their closeout dates. In the future, we will integrate the inbound Paperless Fax Imaging Reporting System with the imaging system currently being developed.

Post-Data Collection Processing

The Computer-Assisted Survey Research Office's (CASRO) re-invention lab addressed re-engineering the processes that traditionally follow data collection. We began this effort to evaluate, procure and/or design survey processing tools to standardize processing systems throughout the U.S. Census Bureau. These systems usually perform similar functions—editing, imputation for missing data, weighting, tabulating, and statistical analysis—but are usually custom-coded for specific surveys on a wide variety of hardware/software platforms. As surveys move to automated data collection and edits are built into the data collection instrument, the U.S. Census Bureau needs to reconsider traditional post-collection processes. To begin re-engineering these processes, the re-invention lab collected high-level requirements for developing common processing tools that survey managers can use in an integrated computing environment (ICE). CASRO used these requirements to develop an ICE white paper that documents the functionality of ICE and an IT architecture to support the environment. As part of this process, the ICE team developed a business data model that forms the basis of the U.S. Census Bureau's corporate metadata repository standards.

ICE will facilitate many of the common activities associated with a census or survey. It will provide a common user interface to a set of generalized planning, processing, retrieval, and analytical tools. The tools will access the U.S. Census Bureau data library that will contain census data and coinciding metadata. Appropriate security measures will protect the information stored in the data library. CASRO is supporting efforts that advance the ICE concept; these efforts include research and planning for an agency metadata system, the Economic program

area's Standard Economic Processing System, and the U.S. Census Bureau's American FactFinder.

Data Dissemination

The Computer-Assisted Survey Research Office (CASRO) is working with other U.S. Census Bureau organizations to design and implement a data dissemination capability that will provide customers with as many options as possible to access census information. Beyond the traditional paper publications, the U.S. Census Bureau provides the option to purchase publications on CD-ROM and to access information through an Internet web-server. The U.S. Census Bureau's priority is to provide the public with attractive products and options for access; CASRO is supporting the development of a Bureau-wide corporate metadata repository that will provide the metadata needed to share data across the organization and disseminate information to external users.

American FactFinder, a subsystem of the Data Access and Dissemination System (DADS), is the overall initiative to offer Internet/Intranet access to the U.S. Census Bureau's data. DADS's initial focus was to provide access to data for Census 2000, the American Community Survey, and the 1997 Economic Census. The Federal Electronic Research and Review Extraction Tool (FERRET) database currently provides Internet access to Current Population Survey data. In the future, the U.S. Census Bureau will integrate demographic survey data as a component of American FactFinder.

Our strategy is to provide one general electronic system for customers to access census and survey data, to create customized products, and to acquire pre-designed products. The Integrated Information Solutions

(IIS) program is an effort to reorganize the U.S. Census Bureau's data creation and delivery systems so that we can deliver "integrated information solutions" to our customers inside and outside the Bureau. IIS is an answer to a customer's question, delivered without the customer's first having to learn how the U.S. Census Bureau, our programs, or our data files are organized. The DADS project was a starting point for the IIS program. In June 1998, a cross-divisional/cross-functional team was established to consider the future of data access and dissemination at the U.S. Census Bureau. The team is using the business process reengineering methodology developed by Canal Bridge Consulting and supported by the CASRO staff. The work of the full team was completed in January 1999; work yet to be completed includes beginning detailed implementation planning with pilot demonstration projects.

The IIS program will eventually incorporate the data delivery component of DADS, along with FERRET and other dissemination vehicles. The IIS program will develop new software and procedural tools to allow U.S. Census Bureau staff to more efficiently develop and disseminate data and help customers. While we hope to get a head start with pilot or demonstration projects (Phase 1 is from February 1999 through September 2000), much of our work and development will have to wait until we have funding (Phase 2, starting October 2000). We will put many of those systems in place within a year or two after that, even as we continue development to meet new challenges (Phase 3, starting in 2002 or 2003).

Our Census Modernization initiative supports all the activities discussed above. Those activities, in turn, support program areas across the U.S. Census Bureau.

2.1.2 Computer-Assisted Survey Information Collection (CASIC) Progress Against Planned Milestones

CASIC Milestones, FY 98					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
PFIRS Transition Plan	06/97	05/98		05/98	Completed.
Establish Process Improvement component	09/97	05/98		05/98	Completed.
Travel Redesign	09/97	01/98		01/98	Completed.
CSAQ Transition Plan	11/97	07/98		07/98	Completed.

CASIC Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Integrated Information Solutions Initial planning	06/98	01/99		01/99	Completed.
Survey Process Improvement Team	06/98	01/99		01/99	Completed.
CSAQ Internet Test for the Library Media Center Survey	04/98	02/99		02/99	Completed.
Establish CATI/CAPI Project Plan	10/98	03/99		03/99	Completed.
CSAQ Internet Test for the Company Organization Survey	04/98	09/99			In progress.

CASIC Milestones, FY 00					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Imaging Pilot for the Annual Survey of Manufactures (ASM)	06/98	12/99			In progress.
IIS Pilot Facilitation	04/99	12/99			In progress.
CSAQ Internet Production for the Library Media Center Surveys	04/99	02/00			In progress.
Imaging Pilot for American Community Survey	06/99	04/00			In progress.
CSAQ Internet Test for ACS	06/99	04/00			In progress.

CASIC Milestones, FY 01

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Workflow Test	01/00	12/00			
Imaging Transition Plan	01/00	12/00			
CSAQ Internet Transition Plan	06/00	06/01			
Computer-Assisted Telephone/Personal Interviewing Graphic User Interface Research	03/98	09/01			

CASIC Milestones, FY 02

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Migration of CATI/CAPI Surveys to Windows (Phase 1)	09/01	09/02			
Migration of Surveys to Imaging (Phase 1)	01/00	09/02			
Migration of Surveys to CSAQ Internet (Phase 1)	01/00	09/02			

CASIC Milestones, FY 03

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Development of Integrated Computing Environment Tools	10/00	09/03			

Progress to date: The Computer-Assisted Survey Research Office (CASRO) completed the following four projects during Fiscal Year 1998:

- implementing the redesigned travel system;
- finishing the transition plans for the Paperless Fax Imaging Reporting System; (continued)

- finishing the transition plans for the Computerized Self-Administered Questionnaire; and
- establishing the Process Improvement Component.

We concluded Internet testing of Computerized Self-Administered Questionnaires in early Fiscal Year 1999. Training the CASRO staff remains an ongoing activity.

2.1.3 Computer-Assisted Survey Information Collection (CASIC) Performance Measures

Fully implementing CASIC will improve performance (through accelerated processes) and reduce labor-intensive activities. It will also lead to lower costs and improved information flow to our customers. Initially, the U.S. Census Bureau will measure performance through Initial Technical Assessments (benchmarks on what other organizations have learned) and prototype testing and evaluations. Once fully implemented, it will measure performance with milestones, schedules and internal status updates assessing our implementation progress. We will improve service quality to its customers and respondents by offering reporting options; we will measure service quality through customer satisfaction surveys as well as changes in response rates and costs. Performance measures will largely be related to specific program areas' applications and will therefore be built into their processes and reported in their sections of the IT plan.

Implementing CASIC will increase the U.S. Census Bureau's capability for additional federal reimbursable work, with the potential benefit of reducing respondent burden. CASIC will enhance the Department of Commerce's ability to respond quickly to information requests from Congress, the President, businesses, academia, and other data users. It will provide an opportunity to broaden the access and use of census data by providing data access tools that support the dissemination of information through the National Information Superhighway.

Some of the typical measures that the program areas could track are:

- reducing data delivery time of data (reduced cycle time); and
- improving data quality (reducing errors) through automation.

2.1.4. Computer-Assisted Survey Information Collection (CASIC) Risks

Before moving any new technology to a production environment, we prepare a transition plan that includes the technology's purpose, cost/benefit component, production environment, resource needs, training needs, funding, divisional responsibilities and schedule. These transition plans discuss the

risks associated with implementing the specific technology. Various customers assume the risks, and the production areas' Information Technology Plans discuss the risks further. Please see section 2.1.5, below, for supporting documents.

2.1.5 Computer-Assisted Survey Information Collection (CASIC) References

CASIC is supported by the following planning documents:

- Budget submission for FY 2001, dated June 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 83-84, 99;
- Integrated Computing Environment White Paper, dated September 19, 1997;
- *Computer-Assisted Survey Information Collection*, Requirement Initiative DR01-9401, approved August 25, 1994, renewed July 6, 1998 through 2002;
- Transition Plans for STAMP, CSAQ, PFIRS, and Metadata; and
- the following Business Plans:
 - Technology: Computerized Self-Administered Questionnaires;
 - Technology: Develop Transition Plan for Paperless FAX Image Reporting System;
 - Technology: Establish a Re-engineering Services Component to CASRO; and
 - Re-engineering: Conduct Process Improvements for Three Prototype Projects.

2.2.1. Detailed Description of Integrated Statistical Laboratory

The Integrated Statistical Laboratory provides a statistical research computing environment with the power and flexibility to handle the following:

- multiple operating systems;
- many users;
- large files and databases;
- a large variety of statistical, analytical, and graphical packages and libraries;
- highly computer-intensive customized programs; and
- the network communications to permit researchers to share information easily and accurately.

Our Sun cluster and PC network systems have evolved continuously over the last ten years with yearly technology upgrades and phased retirement of older systems. These systems are part of the basic computing infrastructure that supports research in the Methodology and Standards Directorate.

The Integrated Statistical Laboratory conducts broad research and development activities in the following areas:

- Automated Coding;
- Cognitive Methods;
- Computer-Assisted Survey Information Collection Technology;
- Editing and Imputation;
- Ethnographic Research;
- Quality Control;
- Questionnaire Design;
- Record Linkage;
- Small Area Estimation;
- Survey Methods Research;
- Survey Sampling;
- Time Series;
- Undercount Research (Accuracy and Coverage Evaluation); and
- Variance Estimation.

Research and prototyping activities for new Computer-Assisted Survey Information Collection technologies are being accomplished within the framework of the Integrated Statistical Laboratory network. These projects include Touchtone Data Entry, Voice Recognition Entry, Large Vocabulary Recognition Entry, and fax reporting and transmission of forms.

Research activities have been established for the use of administrative records for censuses and surveys.

The Integrated Statistical Laboratory network occupies the following parts of Federal Office Building 4:

- part of the 3200 wing (shared with non-U.S. Census Bureau personnel);
- part of the 3100 wing (shared with the Economic Statistical Methods and Programming Division); and
- rooms 3000 and 3101.

Services are provided by two resources: a PC-based Netware server and a cluster of Sun UNIX servers. The Netware server is

running Netware 4.0 and was configured following the IT Directorate Telecommunication Office's guidelines for inclusion in the U.S. Census Bureau's enterprise Network Directory Service tree structure. This server uses the IPX protocol for network communications. The Netware 4.0 server provides basic office automation applications and will become a backup to Netware servers once it is replaced by a Netware 5 system with enhanced disk and memory capabilities.

The UNIX network is a mixture of Sun SPARC and Hewlett Packard devices. These devices use the TCP/IP protocols. The primary UNIX servers are four Sun Ultra 2300 dual 300 MHz processor machines and an older eight 40MHz processor SparcCenter 2000. Users primarily access these servers through desktop PCs and X-terminals. The PCs are almost all Pentium models of various speeds. Thirty-five laser printers are also available, with all but two attached to the network. These devices support 110 people in the Integrated Statistical Laboratory and Administrative Records Research staffs, as well as users in other divisions who collaborate on research projects.

The Integrated Statistical Laboratory network has a mixture of standard Ethernet (10Mbps) and fast Ethernet devices (100Mbps) supported by Cisco 5000 switches in the 3100 and 3200 wing telecommunications closets. These fast Ethernet switches are essential to the architectural design of the system, supporting fast access to data from any of the Sun Ultra 2300s in the cluster as well as fast data access to researchers' desktop computers.

The other 12 UNIX desktop workstations are Sun processors of various models and age (except for one Hewlett Packard 730),

that vary substantially in performance, disk capacity, and display capabilities. These workstations provide CPU and local storage dedicated to individual projects, as well as a high-resolution graphical display.

We also have 20 X-terminals that are a substantially less expensive alternative to workstations, yet provide a highly responsive X-window environment that communicates easily with any VAX or UNIX machine on the network. We use XonNet software on some PC desktops to also provide access to X-windows.

The Integrated Statistical Laboratory UNIX platforms provide programming and soft-

ware development tools, data analysis software, graphical packages for statistics and presentations, and advanced mathematical and statistical software. These software tools are critical parts of nearly every research project, and it is essential for effective research that these be highly capable.

The HP server is an HP730 processor dedicated to supporting the Industry and Occupation coding project, which is shared by the Housing and Household Economic Statistics Division and the Statistical Research Division.

2.2.2 Integrated Statistical Laboratory Progress Against Planned Milestones

Integrated Statistical Laboratory Milestones, FY 98					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/97	09/98			Completed.

Progress to date: Consistent with its open systems philosophy, the Integrated Statistical Laboratory has continued to upgrade its hardware and software resources. The Integrated Statistical Laboratory continues to provide maintenance contracts for both hardware and software.

Integrated Statistical Laboratory Milestones, FY 99					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/98	09/99			In progress.
Integrated Statistical Laboratory Upgrade Storage	10/98	09/99			In progress.

Integrated Statistical Laboratory Milestones, FY 00

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
ISL upgrade Timberwolf tape drives	10/99	12/99			
All PCs to be running Windows NT	10/99	12/99			
Upgrade all PCs to at least the current standard	10/99	09/00			
Upgrade CPUs – ULTRAS	01/00	09/00			
Retire SRD2000	03/00	09/00			

Integrated Statistical Laboratory Milestones, FY 01

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/00	09/01			
Start Gigabit Network	10/00	09/01			
Upgrade storage	10/00	09/01			

Integrated Statistical Laboratory Milestones, FY 02

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/01	09/02			
Continue Gigabit Network	10/01	09/02			
Upgrade CPUs Network	10/01	09/02			

Integrated Statistical Laboratory Milestones, FY 03

Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
Upgrade all PCs to at least the current standard	10/02	09/03			

2.2.3 Integrated Statistical Laboratory Performance Measure

Integrated Statistical Laboratory Performance Measure			
Performance Goals	Performance Measures	Target Performance	Current Performance
Use the systems to start research	% of program areas and others who adapt statistical methods and software	100%	100%

2.2.4 Integrated Statistical Laboratory Risks

There are significant risks to the continued success of the Integrated Statistical Laboratory; these risks are, primarily, loss of skilled systems staff and loss of funding. Another risk is the loss of the Statistical Research Division computer room in Federal Office Building 4. This room holds the crit-

ical Statistical Research Division UNIX servers. No real adequate backup systems exist for these servers, although we could use other U.S. Census Bureau UNIX servers to run critical applications. We will protect backup data by placing it in a safe in the secure area at the Bowie Computer Center.

2.2.5. Integrated Statistical Laboratory References

The Integrated Statistical Laboratory is supported by the following planning documents:

- Budget Submission for FY 2001, dated June 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 82-84, 99;
- *Integrated Statistical Laboratory*, Requirement Initiative SM02-9801, approved April 28, 1998; and
- the Annual Report of the Statistical Research Division.

2.3.1 Detailed Description of Administrative Records Research System

This research program's goal is to expand statistical uses of administrative records to reduce respondent burden and costs, improve coverage of censuses and surveys, and generate new information not otherwise available from surveys or censuses. Before being deemed statistically reliable, administrative records data must undergo extensive research to address data quality, consistency with census definitions, and coverage. This research supports uses of administrative

records for decennial, intercensal estimates, survey and address list applications.

Although the primary purpose of the Administrative Records Research System (ARRS) is to support research, the U.S. Census Bureau is planning several program applications—most notably for Census 2000, Master Address File Updates, and the intercensal estimates programs—that will require ARRS' resources. Census 2000 will use administrative records to conduct an experi-

mental administrative record census (AREX2000) in parallel with the traditional census, then compare the results. The U.S. Census Bureau is considering using administrative records as part of the design for updating the Master Address File after Census 2000. For the intercensal estimates program, the U.S. Census Bureau will acquire, merge, and geocode administrative files to support expanding the estimates to smaller geographic levels.

Data processing requirements for administrative records research are significant. The large size of files received from external sources, particularly the files received from the Internal Revenue Service and the Social Security Administration drives our disk storage and CPU needs. Files currently utilized for research and considered primary candidates for uses are:

- **SSA Numident:** a file of approximately 765 million records, containing transactions for all Social Security Number holders, including name, race, date of birth, and gender;
- **IRS Tax Returns (1040):** a file of approximately 120 million tax returns, containing Social Security Numbers, names, and addresses;
- **IRS Information Returns (W-2s and 1099s):** a file of approximately 1.2 billion records, containing Social Security Numbers, names, and addresses;
- **Medicare:** a file of 35 million individuals receiving Medicare benefits, containing Social Security Numbers, names, addresses, races, dates of birth, and genders;
- **Selective Service:** a file of 13 million registered males, containing Social Security Numbers, names, addresses, and ages;
- **Indian Health Service:** a file of several million individuals receiving health services, containing Social Security Numbers, names, addresses, races, ages, and genders; and
- **Other:** including Housing and Urban Development files, food stamp program files, and some state or local files.

We have essentially completed a full system design. To date, research has used existing systems and has been limited by available disk storage. The logical process we have used has been to load and read the source files, edit them, and produce a standardized format file. We geocode the addresses, merge the files, and unduplicate the records. We then build composite records, establish an address/person file link, and produce a final output file for evaluation.

The principles we employ to create this prototype include the following:

- the design of the initial prototype should not prohibit future incorporation of longitudinal updates;
- the design of the initial prototype should not prohibit future inclusion of data from surveys/censuses (including economic);
- the initial prototype should be easily adaptable to future uses and needs;
- we should purchase, rather than develop, software wherever possible;
- the system will be Y2K compliant;
- the system will be "open" (i.e., POSIX compliant);
- the system will conform to U.S. Census Bureau archival standards; and
- the system should pass through the large Administrative Records file as few times as possible.

To date, research into Decennial and Demographic use of administrative records has been done using existing IT Directorate and Statistical Research Division computer systems. We completed a major upgrade of our existing hardware in the fall of 1998. We upgraded our existing 8400/5 350 VAX Alphaserver and purchased a second VAX Alphaserver. Included in the purchase of these machines was a five-year hardware maintenance contract.

The core Administrative Records Research System processing system now consists of two VAX Alphaserver 8400 5/625 systems; each system has eight CPUs and four GB of main memory. One system (DMCA01) is configured to operate in VMS and has one terabyte of RAID storage. The second system (DMCA01) is configured to operate in UNIX and has two terabytes of RAID storage. Each system also has a TL893 264 cartridge/3 drive DLT tape library. The two

DLT tape libraries can hold a maximum of 18.4 terabytes (native) of data.

The primary reason we have configured these machines under two different operating systems is the difficulty in porting hundreds of FORTRAN programs from DEC VMS to DEC UNIX. We are beginning a gradual phased migration from DEC VMS to DEC UNIX.

2.3.2 Administrative Records Research System Progress Against Planned Milestones

Administrative Records Research System Milestones, FY 98-03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
No milestones to report					

2.3.3 Administrative Records Research System Performance Measures

Administrative Records Research System Performance Measures			
Performance Goals	Performance Measures	Target Performance	Current Performance
Reduce respondent burden, reduce data collection costs and more timely statistics	% research results delivered to customers and policy makers on schedule	100%	N/A

2.3.4 Administrative Records Research System Risks

Our research program would be severely compromised if access to key files, such as the Internal Revenue Service (IRS) Individual Master File or the Information Return Master File, were eliminated. In accordance with our memorandum of understanding with the IRS, they will be performing a site audit in 1999; this site audit includes re-

viewing our core processing systems, our internal customer's systems, and the systems in the Economic program area. If the results of this review are negative, our ability to access IRS data agency-wide could be limited or even eliminated.

The success of our research program is dependent on our ability to hire and retain

skilled programming staff. Because government salaries are not competitive with the private sector, we are hampered in attracting candidates for vacancies.

Finally, because our servers are housed at the Bowie Computer Center, many of the risks that apply to that facility apply to the Administrative Records Research System as well.

2.3.5 Administrative Records Research System References

The Administrative Records Research System is supported by the following planning documents:

- Budget Submission for FY 2001, dated June 1999;
- 1999 Strategic IT Plan, dated December 18, 1998, pages 83-84, 99;
- *Integrated Statistical Laboratory*, Requirement Initiative SM02-9801, approved April 28, 1998; and
- *CASIC Research and Development*, Requirement Initiative CASI-9501, approved August 21, 1995.

3.0 Methodology and Standards Program Area Infrastructure Description

3.1 Detailed Description of Infrastructure System

The Methodology and Standards program area infrastructure consists of the following:

- a Netware server running Netware 4.0;
- 183 Pentium PCs;
- 10 other PCs;
- 33 printers;
- IPX telecommunications protocol; and
- fast Ethernet devices.

Two different technical staffs maintain the Methodology and Standards program area infrastructure. ADNET maintains the Com-

puter-Assisted Survey Research Office's infrastructure; please see the ADNET description in the Enterprise IT Support portion of this Plan. The other Methodology and Standards areas use in-house staff to support their infrastructure; these staffs resolve problems, maintain systems, and upgrade hardware and software. In addition, the Integrated Statistical Laboratory use hardware maintenance services (for both PCs and UNIX) provided by the Small Computer System Repair and Maintenance Staff.

3.2 Methodology and Standards Program Area Infrastructure Progress Against Planned Milestones

Methodology and Standards Infrastructure Milestones, FY 98-03					
Description	Estimated		Actual		Progress to Date
	Start Date	Finish Date	Start Date	Finish Date	
No milestones to report					

3.3 Methodology and Standards Program Area Infrastructure Performance Measures

The Methodology and Standards infrastructure has no performance measures to report.

3.4 Methodology and Standards Program Area Infrastructure Risks

The Statistical Research Division has a single Novell server for its office automation functions; a failure could cause serious problems. We will mitigate this by purchasing a new server and using the current server as a backup device.

3.5 Methodology and Standards Program Area Infrastructure References

The Methodology and Standards program area infrastructure is supported by the following planning documents:

- *Enterprise Printer Resources*, Requirement Initiative PRMAP9701, approved May 13, 1998; and
- *Enterprise Personal Computer Management and Acquisition Plan*, Requirement Initiative PCMAP9601, approved December 5, 1995.